APPENDIX E

Statistical Analysis Approach

APPENDIX E

Statistical Analysis Approach

Modeling Mean Emissions

The log-linear model was chosen to describe the data. For each species, emissions were modeled versus mileage, m_{ij}, as

$$\log(E_{ij}) = \alpha + \beta \log(m_{ij}) + v_i + \delta_i \log(m_{ij}) + \epsilon_{ij}.$$

where E_{ij} is the measured emissions in (g/mi) from the *j*th test on the *i*th vehicle. The terms α (intercept) and β (slope) represent the systematic fleet-specific effects. The terms v_i and δ_i represent vehicle-specific deviations from the fleet-specific effects. The final term ϵ_{ij} represents testing variability that may include variations in test procedures and chemical analyses. Vehicle and test variability terms, v_i, δ_i and ϵ_{ij}, are assumed to be normally distributed.

Because the model selected for emissions is lognormal (i.e., $\log(E_{ij})$ has a normal distribution), the average emissions within a fleet is

$$E(emissions) = \exp\left(\mu + \frac{\sigma^2}{2}\right).$$

The mean, μ, on the log scale depends on mileage and is expressed in terms of the model parameters as

$$\mu = \alpha + \beta \log(m) \tag{E-1}$$

where m is the mileage/10,000. The variance, σ^2, is expressed in terms of vehicle and test variance as

$$\sigma^2 = \sigma^2_{vehicle} + \sigma^2_{test}.$$

Thus, vehicle variability can be expressed in terms of the modelled random effects as follows:

$$\begin{aligned} \sigma^2_{vehicle} &= var(\nu_i + \delta_i \log(m)) \\ &= var(\nu_i) + \log(m)^2 \, var(\delta_i) + 2\log(m) \, cov(\nu_i, \delta_i) \end{aligned} \tag{E-2}$$

It was assumed that ν_i and δ_i are distributed according to a multivariate normal distribution as shown below.

$$\begin{bmatrix} \nu_i \\ \delta_i \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} var(\nu_i) & cov(\nu_i,\delta_i) \\ cov(\nu_i,\delta_i) & var(\delta_i) \end{bmatrix} \right) .$$

The covariance term $cov(\nu_i, \delta_i)$ is included to allow for a possible correlation between a vehicle's mean emissions and its rate of degradation with mileage. However, statistical tests suggested very little correlation between these factors. Therefore it was assumed that these factors were independent. Variance estimates were pooled across fleets, except where significant differences in the variability of the transformed levels among fleets were observed. Table E-1 summarizes these instances.

Table E-1. Compounds for Which Variability Estimates Depended on Fleet

Compound	Pool 1	Pool 2
CO, NO_x, benzene	RFG, RF-A	CNG, propane gas, M-85
Formaldehyde	RFG, RF-A, M-85	CNG, propane gas
1,3-butadiene	RFG, RF-A, M-85, Ford propane gas	CNG, Chevrolet propane gas[a]
N_2O[b]	Chevrolet RFG	All other fleets

[a] The CNG and Chevrolet propane gas vehicles had no measurable emissions of 1,3-butadiene. To avoid underestimating variability in emissions levels, these fleets were not included when fitting the model. For these fleets, mean emissions were estimated as zero.

[b] There was insufficient data on emissions of N_2O from the unleaded vehicles to fit models versus mileage.

Estimated Percent Increase for Additional 10,000 Miles

In Table 7, estimates were presented of the average percent increase in emissions of CO, NMOG, and NO_x from 10,000 to 20,000 miles. These estimates were based on the fitted models. Based on an analysis simlar to that described above for estimating mean emissions, the estimated increase from mileage m to m+Δ (measured in 10,000 mi) is:

$$100 * \left(\frac{\exp\left(\alpha + \beta \ \log(m+\Delta) + \dfrac{var(\nu) + \log(m+\Delta)^2 var(\delta) + \sigma^2_{test}}{2}\right)}{\exp\left(\alpha + \beta \ \log(m) + \dfrac{var(\nu) + (\log(m))^2 var(\delta) + \sigma^2_{test}}{2}\right)} - 1 \right)$$

$$= 100 \ * \left(\exp\left(\beta \ \log\left(\frac{m+\Delta}{m}\right) + \frac{Var(\delta)}{2}\left[(\log(m+\Delta)^2) - (\log(m))^2\right]\right) - 1\right)$$

From 10,000 to 20,000 miles, this increase is

$$100 \ * \left(\exp\left(\beta \ \log \ 2 + \frac{Var(\delta)}{2}(\log \ 2)^2 \right) - 1 \right) \ .$$

Comparing Alternative Fuel Emissions With a Control

In making comparisons between emissions of alternative fuel (alt) vehicles and control (ctrl) vehicles, ratios of mean emissions were estimated. Thus, the parameter of interest is:

$$\frac{\exp\left(\mu_{alt} + \dfrac{\sigma^2_{alt}}{2}\right)}{\exp\left(\mu_{ctrl} + \dfrac{\sigma^2_{ctrl}}{2}\right)} = \exp\left(\mu_{alt} - \mu_{ctrl} + \frac{\sigma^2_{alt}}{2} - \frac{\sigma^2_{ctrl}}{2}\right) \ . \tag{E-3}$$

The effect of mileage, m, on μ_{alt}, μ_{ctrl}, σ^2_{alt}, and σ^2_{ctrl} is suppressed from the notation, but was included in the computations. This effect was illustrated in equations E-1 and E-2.

In most cases, the variance components were very similar, in which case the latter two terms on the right hand side of equation E-3 canceled each other, and confidence intervals were derived based on the standard error of $\mu_{alt} - \mu_{ctrl}$. However, for the compounds indicated in Table E-1, differences in the estimated variance components between the alternative and control fueled vehicles could not be ignored. In these cases, the estimated uncertainty in estimating $\sigma^2_{alt} - \sigma^2_{ctrl}$ was also taken into account.

Modeling Proportions

Some of the analyses required modeling of proportions such as the determination of the percent contribution of light- and mid-range hydrocarbons, alcohols, and carbonyls to total NMOG and total ozone. The percent contribution of propane to total NMOG emissions from propane gas vehicles, and methanol from M-85 vehicles was also modeled. For these ratios, observed variability was reasonably constant, so the data were not transformed for analysis. The mixed models, including mileage and vehicle effects, were fit directly to the observed proportions. For these responses, mileage was found to be a significant factor only in the contribution of propane to total NMOG from the propane gas vehicles. The only other caveat is that a significant difference was observed in the contribution to total NMOG from RFG vehicles between the first round of emissions tests and the second and third rounds of emissions tests. This is discussed in the Ozone Reactivity section of this report.

Miscellaneous Modeling

Modeling relative specific ozone reactivity adjustment factors required two steps. First, the measured specific ozone reactivity (SOR: observed ratio of total computed ozone reactivity to total NMOG) was modeled as a linear function of mileage. The impact of mileage was significant for several fleets (all RF-A, Chevrolet and Dodge RFG, and Chevrolet CNG). This provided estimates of SOR for each fleet with confidence intervals as a function of mileage.

Fieller's theorem (see Reference 2) was then used to provide confidence intervals for the relative specific ozone reactivity for each alternative fuel (i.e., the ratio of mean SOR for each alternative fuel to the mean SOR for its respective control fleet). Because of the dependence on mileage, this needed to be performed at multiple mileages.

Formaldehyde Emissions by Bag

Measured emissions of formaldehyde from M-85 vehicles by bag were modeled linearly as a function of mileage. No log transformation to either the response or mileage was deemed necessary.

SHED Tests

Due to the small number of evaporative emissions tests performed, simple averages and standard deviations were calculated for each of the evaporative emissions considered. No attempt was made to separate vehicle-to-vehicle variability from replicate test variability.

1. Searle, S.R., Linear Models, Wiley and Sons, New York, 1976.

2. Kotz, S., and Johnson, N.L., Encyclopedia of Statistical Sciences, Vol. 3, Wiley and Sons, 1983.